AD A043900

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 6	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) PRELIMINARY STEPS IN THE DESIGN OF OPTICAL CHARACTER RECOGNITION ALGORITHMS		5. TYPE OF REPORT & PERIOD COVERED Reprint
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) R. J. Shillman and T. T. Babcock		8. CONTRACT OR GRANT NUMBER(s) DAAB07-76-C-1400, NSF-ENG-74-24344
9. PERFORMING ORGANIZATION NAME AND ADDRESS Research Laboratory of Electronics Massachusetts Institute of Technology Cambridge, Massachusetts 02139		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office P. O. Box 12211 Research Triangle Park North Carolina, 27709		12. REPORT DATE August 1977
		13. NUMBER OF PAGES 7
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

DDC
RECEIVED
SEP 7 1977
B

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Optical Character Recognition
Physical-to-Functional Rules

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Previous publications have described a theory and techniques for designing optical character recognition algorithms using psychophysical methodology, focusing on the methodology used to determine Physical to Functional Rules (PFRs). The present paper describes important preliminary steps that precede the psychological experiments, including the choice of an appropriate letter pair for study, the formation of a hypothesis regarding relevant attributes that distinguish the letters of the pair, and the design of stimuli for the subsequent psychophysical experiments. These steps are described in the context of the problem of designing an algorithm for recognizing handprinted 2's from handprinted Z's.

PRELIMINARY STEPS IN THE DESIGN OF OPTICAL CHARACTER RECOGNITION ALGORITHMS

Robert J. Shillman*
Robert T. Babcock

Research Laboratory of Electronics
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ABSTRACT

Blesser et al. [1] and Shillman et al. [2] have described a theory and techniques for designing optical character recognition algorithms using psychophysical methodology. The previous publications have focused on the methodology used to determine Physical to Functional Rules (PFRs) whereas the present paper describes the important preliminary steps that precede the psychological experiments. These steps include the choice of an appropriate letter pair for study, the formation of a hypothesis regarding relevant attributes that distinguish the letters of the pair, and the design of stimuli for the subsequent psychophysical experiments. These steps are described in the context of the problem of designing an algorithm for recognizing handprinted 2's from handprinted Z's.

I. INTRODUCTION

Although the basic idea of having a machine recognize printed characters has been seriously discussed for over 40 years, the ultimate machine that would recognize unconstrained handprinted characters at error rates comparable to human performance has yet to be built. The experience of the past has shown us that Optical Character Recognition (OCR) is indeed a difficult problem; the stages occurring between the insertion of a document and the recognition of the letters on that document involve complex aspects of mechanics, optics, electronics and psychophysics. Each stage has its associated problems along with constrained solutions that may affect the design and performance of later stages. For example, if fixed thresholds are set for character binarization, then variations in printing blackness or photodetector sensitivity will degrade machine performance independently of the sophistication of the subsequent recognition stage. This indicates that there must be some degree of interaction between various stages in order to optimize performance.

Thus we see that there are many stages in the OCR process, and that coupling must eventually be present between the various stages in order to optimize overall machine performance. We have, however, focused our efforts primarily on the design and optimization of the recognition algorithms with the expectation that an understanding of the issues involved in this final stage of the OCR process will yield insight into the specification and design of the preceding stages.

II. DIFFICULT CASES

If each letter and numeral occurred in only a few well-defined shapes, then machine recognition of these characters would be trivial. Unconstrained handprinted characters, however, occur in an infinite variety and are therefore not easily recognized by machine. There are different approaches one might take when first faced with the issue of designing an OCR algorithm; for example, one might focus on the majority of A's in a given data set and attempt to describe them in a formal way. This approach is problematic in that the descriptions arrived at for two different letters may in the end be identical (for example, all A's might be described as having a closure at the top and two descending legs, a description that also applies to all R's). These cases must then be disambiguated through modification of feature weights or by introduction of additional features. Rather than save this step for last, our approach has been to focus directly on these problem cases rather than on typical letter shapes. Once the difficult cases are analyzed, the remaining cases are expected to be straightforward (e.g., there is little problem in devising a simple algorithm for distinguishing M's from T's). Previous publications [3] have detailed the psychophysical techniques employed for studying the problem cases such as V's and Y's, U's and V's, etc. The approach has recently led to a successful low-error-rate computer algorithm for the recognition of unconstrained handprinted U's from V's, which are among the most difficult characters for both humans and machines to identify correctly [4].

III. IN SEARCH OF PHYSICAL TO FUNCTIONAL RULES

The PFR

A Physical to Functional Rule (PFR) is a mapping from the physical domain, which is the graphical image of a character, to the functional domain from which the character's label can be determined. The only PFR that has been thoroughly examined, validated, and reported is for the attribute LEG which distinguishes between the letters of the pairs (called confusion pairs) shown in Fig. 1.

Fig. 1. Characters distinguished by the attribute LEG.

In a neutral context, the PFR for LEG is given by:

$$\text{Functional LEG: } \frac{\ell_1}{L} \begin{matrix} > \\ < \end{matrix} \; 0.17 \quad \begin{matrix} \text{Present} \\ \\ \text{Not present} \end{matrix}$$

where ℓ_1 = length of physical leg, and L = length of line of which the leg is a part. This rule indicates

that if the length of the physical leg is greater than 17% of the length of the stroke of which it is a part, then the characters would most likely be labeled as Y, F, P, H and A.

The remainder of this paper describes the important preliminary steps that precede the psychophysical experiments conducted to determine PFRs such as the one we have given. These steps are: choosing an appropriate letter-pair for study, forming an initial hypothesis regarding relevant attributes that distinguish the letters of the pair, and designing the stimuli for a psychophysical experiment to test the hypothesis and generate a PFR.

Choosing a Confusion Pair

The first step in the search for a PFR is to choose a pair of characters that are often or easily confused with each other. Neisser and Weene [5] studied human recognition of handprinted characters, and showed, for example, that many handprinted 2's are misrecognized as handprinted Z's. After the likelihood of occurrence of various confusion pairs has been determined, it may be noted that the potential confusion in a number of pairs appears to result from a common cause, and hence the study of a particular pair may yield insight into other confusion pairs; for example, information on 2-Z discrimination may be useful in understanding the confusion pairs: U-V, S-5, and perhaps in discriminating script n (n) from handprinted M (m).

Initial Hypothesis Formation

When handprinted, the numeral 2 occurs in a large variety of forms, as opposed to the letter Z, which has one basic form. The next step is to examine those 2's that lie near the 2-Z interletter boundary in an attempt to gain insight into the relevant attributes. We can focus on boundary cases by constructing an interletter trajectory from the archetypal shape of one letter to the archetypal shape of the other letter of the confusion pair. Although generally straightforward, this process is complicated in the 2-Z case by the fact that although there is one accepted Z archetype, there are a number of different "standard" forms for the handprinted numeral 2 (for example, 2 , 2 and 2).

Although these forms may have all evolved from one archetype, several forms now occur with sufficient frequency to be considered archetypal [6]. In his very thorough study of numerals, Wright [6] suggests that the numeral 2 can be studied by segmenting it into four regions: the turn, base, stem and head. Each of these segments can occur in a variety of forms summarized descriptively in Fig. 2, and examined in the following text.

a. The turn

For the purpose of 2-Z recognition, the turn can be categorized as looped or simple. This categorization is useful because a looped turn is extremely infrequent in handprinted samples of the letter Z and does not occur near the 2-Z interletter boundary, and is therefore not helpful in defining the recognition boundary between 2's and Z's. This does not mean that this portion of the character would not be used in a recognition algorithm; on the contrary, the presence of a looped turn is a strong indication that the character is a 2 in much the same way that a horizontal crossbar is a strong indicator of Z-ness; these embellishments could be searched for and utilized in the initial stages of the OCR decision process.

b. The base

Although Fig. 2 indicates that there may be four regions where information about the character's label may be located, the base appears not to contain any information relevant to 2-Z discrimination. Figure 3 shows that the base alone does not determine letter label: for any chosen base, the character can be completed so as to become either a 2 or a Z. A more
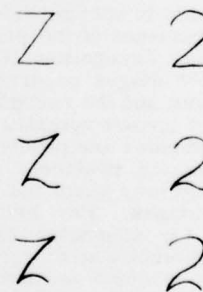

Fig. 3. Negligible effect of the base on character label.

formal method of proving that the base does not contain a Functional Attribute is to note that it does not satisfy the primary definition of this attribute which requires that it must be possible to alter a character's identity along a trajectory by varying the functional attribute. In this case, the test requires that an interletter 2-Z trajectory exist along which only the base is varied; since no such trajectory has been found, the base is ruled out as an informative region. For simplicity, therefore, we shall initially study characters having flat bases and simple turns.

c. The stem

The three types of stems are doubly curved, simply curved, and straight. Figure 4 shows that both 2's and Z's can be constructed to have any of the three possible stems.
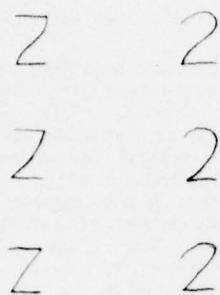

Fig. 2. Regions of the character 2.

Fig. 4. Negligible effect of the stem on character label.

Although the Z's with doubly curved or simply curved stems are not very good, we see that the stem by itself does not determine letter label. It appears that the doubly curved stem is least Z-like, whereas the straight stem is most Z-like. Because there is more than one archetypal 2 shape, the fact that a doubly curved stem is least Z-like does not necessarily mean that it is most 2-like; in fact, all three types of stems occur with roughly the same frequency in hand-printed samples of the numeral 2 [7].
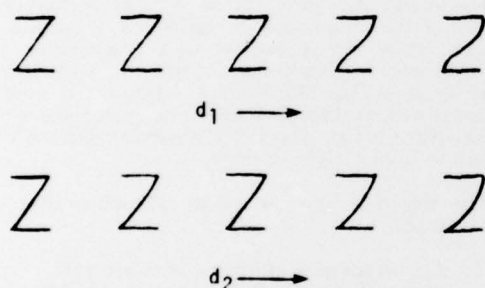

Fig. 5. Effects of increasing double curvature along d_1 and single curvature along d_2.

Fig. 6. Representation of the 2-Z space along four dimensions.

Figure 5 shows two trajectories along which the stem is varied; along dimension d_1 the stem becomes increasingly simply curved. The postulated decrease in goodness [8] along both trajectories is hypothesized to be due to two independent factors; the major decrease is due simply to the degradation of the character (movement away from the central portion of the Z space) and only a small portion is due to the influence of the numeral 2 (movement toward the 2 space). This effect is illustrated in Fig. 6 which shows both trajectories of Fig. 5 replotted on the 2-Z space. Along either of these two trajectories, the characters clearly become much less Z-like and slightly more 2-like; this is illustrated by the movement away from the center of the Z space and only slowly toward the 2 space.

d. The head

For the purpose of 2-Z discrimination, the five heads described by Wright can be categorized as either marked (containing loops or spurs) or simple. As in the case of the looped turn, the presence of a marked head is infrequent in the letter Z and could thus be used as an initial test for character label. Focusing on simple heads, it can be noted that there are two regions that contribute to the perceived curvature, or roundness, of the head: the curvature at the
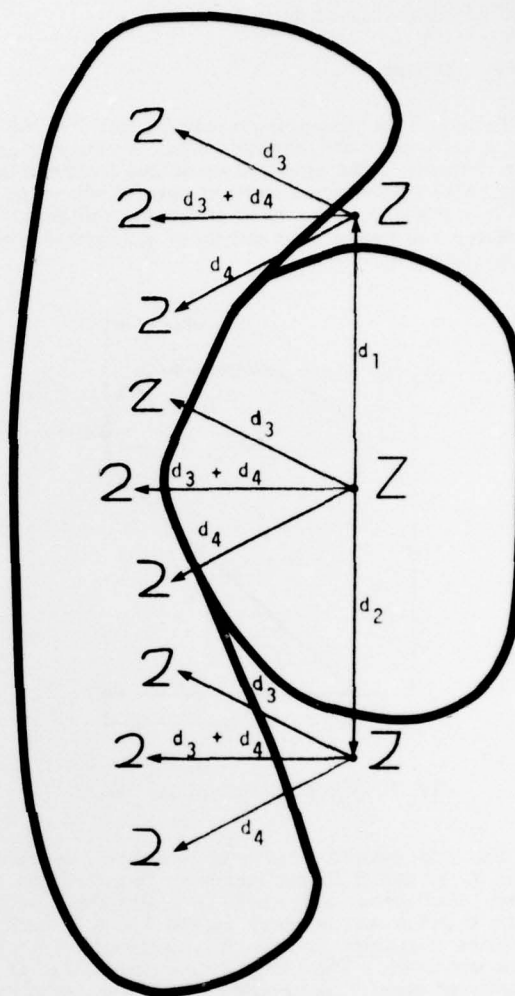
start of the stroke (in the upper left of the character) and the region of curvature in the upper right corner. Figure 6 shows variations along both these dimensions, d_3 and d_4, respectively; although the stem apparently has some effect on letter label, the effect along either d_1 or d_2 is thought (and is pictured) to be small compared to the effects along d_3 and d_4; this indicates that the major change in identity is due to changes in the head of the character. Figure 6 also contains a trajectory directly from the central portion of the Z space to the 2 space along $d_3 + d_4$.

The hypothesized 2-Z space shown in Fig. 6 indicates the complexity of the 2-Z problem which is brought about, in large part, by the presence of multiple archetypal shapes of the numeral 2. It appears that there are at least four physical dimensions that are involved in 2-Z discrimination. One can traverse an interletter trajectory from an archetypal Z to an archetypal 2 via $d_1 + d_3 + d_4$, or $d_2 + d_3 + d_4$, or via $d_3 + d_4$. Since the last trajectory is involved in each of the other two and since its effect is thought to be major, it seems appropriate to investigate initially

the dimensions d_3 and d_4.

Stimulus Design

Based on the foregoing analysis, the 2's initially chosen for study have simple heads, straight stems, and flat bases. The physical variables involved in the constructed trajectories were chosen to affect the perceived curvature of the head of the character. Each character has two curved and three straight segments, as shown in Fig. 7.
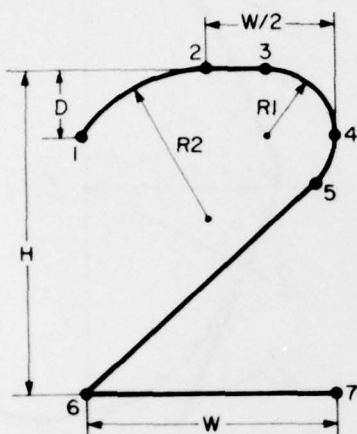

Fig. 7. Construction of the stimuli.

The two curved segments are arcs tangent at points 2, 3, and 5 to the corresponding straight segments. Horizontal alignment is maintained between points 1 and 6 and between points 4 and 7 such that the entire character just fits in a rectangle of constant height and width. The two physical variables, D, the distance of point 1 below the top, and R1, the radius of the second arc, are varied while point 2 is kept centered horizontally.

For given values of D, R2 is found to construct an arc from point 1 to point 2. For given values of R1, point 5 is found to satisfy tangency and construct an arc from point 3 to point 5. Figure 8 shows a two-dimensional 2-Z trajectory plotted on a Calcomp 563 line plotter with a .3 mm Mars technical pen. R1 varies linearly along the horizontal and D varies linearly along the vertical.
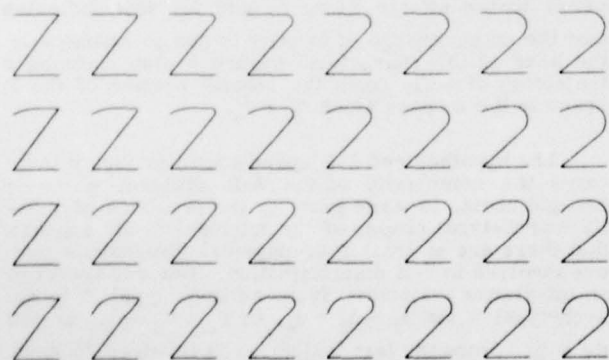

Fig. 8. A two-dimensional 2-Z trajectory.

IV. DISCUSSION

Figure 8 indicates that both physical variables chosen do influence labeling; it cannot yet be determined, however, whether there are two functional attributes involved, each of which takes on as its argument one of the physical variables, or whether there is one functional attribute which takes on both variables as arguments. It does appear that the curvature in the right-hand corner of the head has somewhat greater importance; Fig. 9 shows that if there is an angular bend in the right corner, then no amount of curvature at the start of the stroke and no amount of 2-ness in the stem or base of the character will be sufficient to force the character to be a 2. Note that
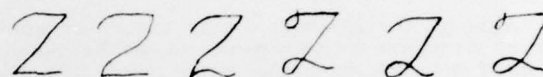

Fig. 9. Dominant effect of an angular bend in the top right corner.

even the addition of a strong 2 indicator, a looped turn, is not sufficient to create a 2 if the upper corner is angular. Studying the trajectories in Fig. 8 it appears that the issue of segmentation [8] may be involved; if the head of the character is composed of one functional segment, then the character is a 2 whereas if it is composed of two functional segments, then it is a Z. It may be that the PFR determining the number of functional segments will take on as its arguments both *the curvature at the start* of the stroke and the curvature in the upper right corner.

The working hypotheses developed in this paper are listed below:

(1) 2-Z discrimination is complicated by the presence of multiple archetypes of the numeral 2.
(2) There are a number of features involved in 2-Z discrimination; in decreasing order of importance they are: (a) the curvature in the upper right-hand corner of the head, (b) the curvature at the start of the stroke, and (c) the nature of the stem, whether straight, simply curved, or doubly curved.
(3) The base is not an informative region for 2-Z discrimination.
(4) Embellishments such as a horizontal cross-bar in the Z or marked head and/or a looped turn in the 2 are strong clues for determining identity and can possibly be treated separately from the features described in (2).
(5) In the absence of embellishments, it may be that the character is a 2 if the head is composed of one functional segment and a Z if it is composed of two functional segments. The PFR determining the number of functional segments may be a function of the three features listed in (2).

We are now undertaking psychophysical experiments to test these hypotheses.

References

1. Blesser, B., R. Shillman, T. Kuklinski, C. Cox, M. Eden, and J. Ventura, "A Theoretical Approach for Character Recognition Based on Phenomenological Attributes," International Journal of Man-Machine Studies, 6, 701-714 (1974).

2. Shillman, R. J., T. T. Kuklinski, and B. A. Blesser, "Psychophysical Techniques for Investigating the Distinctive Features of Letters," International Journal of Man-Machine Studies, 8, 195-205 (1976).

3. Blesser, B., T. Kuklinski, and R. Shillman, "Empirical Tests for Feature Selection Based on a Theory of Human Character Recognition," Pattern Recognition, 8, 77-85 (1976).

4. Suen, C. Y., and R. J. Shillman, "Low Error Rate Optical Character Recognition of Unconstrained Handprinted Letters Based on a Model of Human Perception," to appear in IEEE Transactions on Systems, Man and Cybernetics, June 1977, pp. 73-75.

5. Neisser, U., and P. Weene, "A Note on Human Recognition of Hand-Printed Characters," Information and Control, 3, 191-196 (1960).

6. Wright, G. G. Neill, The Writing of Arabic Numerals (University of London Press, Ltd., London, 1952).

7. Ibid., p. 27.

8. Shillman, R. J., "Character Recognition Based on Phenomenological Attributes: Theory and Methods." Unpublished Ph.D. Thesis, Department of Electrical Engineering, M.I.T., Cambridge, Massachusetts, 1974.